

HUMAN PERFORMANCE MODELS OF PILOT BEHAVIOR

David C. Foyle, NASA Ames Research Center; Moffett Field, CA
Becky L. Hooey, San Jose State University/NASA ARC; Moffett Field, CA
Michael D. Byrne, Rice University; Houston, TX
Kevin M. Corker, San Jose State University; San Jose, CA
Stephen Deutsch, BBN Technologies; Cambridge, MA
Christian Lebiere, Micro Analysis and Design; Boulder, CO
Ken Leiden, Micro Analysis and Design; Boulder, CO
Christopher D. Wickens, University of Illinois; Savoy, IL

Five modeling teams from industry and academia were chosen by the NASA Aviation Safety and Security Program to develop human performance models (HPM) of pilots performing taxi operations and runway instrument approaches with and without advanced displays. One representative from each team will serve as a panelist to discuss their team's model architecture, augmentations and advancements to HPMs, and aviation-safety related lessons learned. Panelists will discuss how modeling results are influenced by a model's architecture and structure, the role of the external environment, specific modeling advances and future directions and challenges for human performance modeling in aviation.

INTRODUCTION

More than two-thirds of all aircraft accidents are attributed to pilot error. Identifying when equipment and procedures do not fully support the operational needs of pilots is critical to reducing error and improving flight safety (Leiden, Keller, & French, 2001). This becomes especially relevant in the development of new flight deck technologies that have traditionally followed a design process more focused on component functionality and technical performance than pilot usage and operability. To help counter this bias and to better understand the potential for human error associated with the deployment of new and complex systems, advanced tools are needed for predicting pilot performance in real-world operational environments. Serious piloting errors and accidents are rare events and the low-probability of occurrence makes the study of pilot error difficult to investigate in the field and in the laboratory. These errors characteristically result from a complex interaction between unusual circumstances, subtle "latent" flaws in system design and procedures, and limitations and biases in human performance. This can lead to the fielding of equipment that puts flight safety at risk, particularly when operated in a manner or under circumstances that may not have been envisioned or tested.

Human performance modeling, when combined with nominal and off-nominal scenario human-in-the-loop testing, provides a complementary technique to develop systems and procedures that are tailored to the pilot's tasks, capabilities, and limitations (Leiden, Laughery, Keller, French, Warwick, & Wood, 2001). Because of its fast-time nature, human performance modeling is a powerful technique to uncover "latent design flaws" -- in which a system contains a design flaw that may induce pilot error only under some low-probability confluence of precursors, conditions and events. Human performance modeling also offers a powerful technique to examine human

interactions with existing and proposed aviation systems across an unlimited range of possible operating conditions. It provides a flexible and economical way to manipulate aspects of the task-environment, the equipment and procedures, and the human for simulation analyses. In particular, modeling and simulation analyses can suggest the nature of likely pilot errors, as well as highlight precursor conditions to error such as high levels of memory demand, mounting time pressure and workload, attentional tunneling or distraction, and deteriorating situation awareness. Fast-time simulation permits the generation of very large sample sizes from which low-rate-of-occurrence events are more likely to be revealed. Additionally, this can be done early in the design cycle, without the need to fabricate expensive prototype hardware.

Five modeling teams from industry and academia were chosen by the NASA Aviation Safety and Security Program to develop human performance models (HPMs) that address two problems in the aviation domain that have significant implications for aviation safety, and that are representative of general classes of problems faced by the aviation industry today. First, modeling teams addressed a current-day aviation problem within the realm of surface operations safety, by identifying causal factors of navigation errors and potential error mitigations (procedural, technical or operational). The notion of understanding causal factors of human error, and the importance of being able to predict when human operators might be vulnerable to error, and predicting which potential mitigating strategies might be successful is pervasive throughout every phase of flight. Second, the modeling teams modeled pilot performance during the approach and landing phases of flight for both a baseline configuration representing today's glass cockpit and a configuration that also included a Synthetic Vision System (SVS). The research issues inherent in this problem are common to the design, development and integration of any advanced cockpit display technology.

Human-In-The-Loop Studies

Effective HPMs require extensive understanding of the task and the domain environment in order to produce valid and meaningful results. HPMs are most informative when supported by empirical data derived from laboratory studies, human-in-the-loop (HITL) simulations, and field studies. To enable model development, information was provided to the modeling teams including task analyses and objective data and subjective ratings from two HITL simulations that were conducted at NASA Ames Research Center. The HITL data were used in two different ways in this project. In some cases the data were used by the modeling teams to populate and develop their models, and in other cases, the data were used to validate the model output.

Airport Surface Operations. A high-fidelity surface operations HITL simulation was conducted to understand the factors that contribute to taxiway navigation errors and potential mitigating solutions (Hooey, Foyle, & Andre, 2001). The simulation, conducted in NASA Ames Research Center's high-fidelity, glass cockpit Advanced Concepts Flight Simulator (ACFS), compared taxi performance under current-day baseline operations with a prototypical cockpit display system called the Taxiway Navigation and Situation Awareness (T-NASA) system which is comprised of an electronic moving map (EMM), a head-up display (HUD), and auditory alerts and warnings. The simulation trials required pilots to land and taxi to the gate following an ATC-issued taxi clearance. All trials were conducted in low visibility at a high-fidelity rendering of Chicago O'Hare airport. The study included common taxi scenarios including hold short instructions and route amendments as well as off-nominal taxi events that represented failures or errors in the system. In current-day

operation (baseline) trials, pilots made navigation errors on approximately 20% of the trials, while these errors were eliminated with T-NASA. NASA provided the HPM teams with data including taxi speed, navigation errors, intra-cockpit communications, pilot-ATC communications, workload, and situation awareness to enable the teams to develop models of the pilot tasks and taxi scenarios necessary to predict taxiway navigation errors.

Synthetic Vision Systems (SVS). A part-task HITL simulation study was conducted to investigate the effect of new synthetic vision systems (SVS) on pilot performance, visual attention, and crew roles and procedures during low-visibility instrument approaches (Goodman, Hooey, Foyle, & Wilson, 2003). The HPM teams were provided with a cognitive task analysis of the approach phase of flight and human performance data including eye movements, communications, and control panel responses from the NASA part-task simulation of instrument approaches with and without the SVS. Events such as no visibility at decision altitude (forcing a missed approach/go-around), a misalignment between instruments and the out-the-window view, and a late runway reassignment, were included in the scenarios to yield a robust set of human performance data.

OVERVIEW OF HPM APPROACHES

From an initial review of past efforts in cognitive modeling, it was recognized that no single modeling architecture or framework had the scope to address the full range of interacting and competing factors driving human actions in dynamic, complex environments (Leiden, Laughery, Keller, French, Warwick, & Wood, 2001). As a consequence, the decision was made to develop and expand multiple modeling efforts to extend the current state of the art within a number of HPM tools. Five modeling frameworks were selected based on a peer-reviewed process with selection criteria including model theory, scope, maturity, and validation as well as the background and expertise of the respective research team.

In Phase 1, each team modeled some aspect of the airport surface operations domain problem with an emphasis on replicating or predicting pilot error. In Phase 2, each team built on their existing model capabilities to address issues relating to SVS design and integration. The approach and specific research questions were left to the discretion of the modeling teams yielding diverse models with a demonstrated capability of answering a variety of important aviation domain questions. In the sections that follow, each modeling team will briefly describe their model architecture, augmentations to the models, and significant findings for the aviation community.

Attention-Situation Awareness (A-SA)

C.D. Wickens (Panelist) and J. McCarley

The Attention-Situation Awareness (A-SA) model has two components. The first (attention, A) describes the way in which three factors of the visual environment – the *salience*, *expectancy* and **value** of events – drive attention allocation, as this allocation is inhibited by a fourth factor, the **effort** required to scan between information sources within that environment. All four factors can be quantified, to make ordinal predictions of the degree to which areas of interest will be fixated as assessed by visual scanning (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). Such visual input supports the second component: situation awareness (SA), or understanding of the current and

future state of the aircraft. SA is based on a belief updating module, such that SA is updated any time a new piece of information is encountered. The pieces are weighted based on the value of the information to the task such that the current value of SA is increased or decreased. SA decays when no new pieces of information are encountered. The model has both an analytic and a real-time dynamic version. The model augments classical optimal scanning models to include the design-layout related factors of salience and effort, to accommodate auditory channels, to consider circumstances related to task priority and those in which there is a one-to-many mapping between fixation areas and tasks.

The surface operations simulation data were used to exercise the model, which predicted both intersections where errors were likely because of degraded SA, and the benefits of the T-NASA display in mitigating those errors. The SVS simulation data from NASA Ames, as well as SVS data collected at the University of Illinois were used to validate the attention allocation component of the A-SA model as inferred from visual scanning measures. The latter data, involving a low altitude flight through a terrain-challenged environment, revealed that the model accounted for an average of 85% of the variance in scanning behavior across the five areas of interest in the cockpit, for eight pilots (Wickens, McCarley, Alexander, Thomas, Ambinder, & Zheng, 2005). Model fit was not improved by considering the inhibiting role of effort, above and beyond that of expectancy and value, suggesting that pilots were quite optimal. Furthermore, greatest deviations from model predictions were shown by those who suffered greater decrements in flight path tracking and hazard detection performance, thereby suggesting that variations in optimality of attention allocation translated to variations in performance.

ACT-R Version 5.0

M.D. Byrne (Panelist), A. Kirlik and M.D. Fleetwood

Our models are detailed closed-loop, pilot-displays-aircraft system models. That is, for both tasks, the ACT-R model of the pilot was connected to an executable model of the aircraft and relevant visual environment and both models run in real time. ACT-R provides a platform to support modeling at a fine temporal detail; the output of an ACT-R model is a time-stamped series of fairly primitive behaviors, down to the individual saccade. ACT-R models are also knowledge-intensive and require extensive task analysis and consultation with subject matter experts to supply the model with the knowledge necessary to actually execute the task. Taxiing and autopilot-guided instrument approach are, in fact, substantially different tasks, and the two models overlap in the general approach taken and the kind of outputs provided but less so elsewhere.

The ACT-R cognitive architecture was originally designed to model the results of laboratory psychology experiments, which typically consist of simple tasks requiring little knowledge and performed in restricted environments. While ACT-R has an extensive track record of being successful in such domains, how ACT-R would “scale up” to aviation-relevant tasks was not entirely clear for both technical (e.g., software integration) and theoretical (e.g., could we represent environmentally-based constraints appropriately in this formalism) reasons. We thus conclude that ACT-R is indeed a viable option for serious HPM research (Byrne & Kirlik, 2005).

Cognitive modeling efforts have traditionally focused on “in the head” cognition, but one of the important lessons learned in these modeling projects is the importance of a high-fidelity

representation of the environment; both of our models are highly sensitive to the structure of the environment. The taxiing model, for example, is very sensitive to the fact that the physical layout of Chicago's O'Hare airport generates taxi routes that are systematically different from typical taxi routes at other airports. This, in turn, contributes to errors in taxiing. The SVS modeling work revealed how a new display with information that is redundant with other cockpit instrumentation has a substantial impact on pilot scanning behavior, even in phases of flight where such changes were not the system designers' intent.

Air-MIDAS

K.M. Corker (Panelist)

Human performance models were developed and applied to surface and flight operations in order to predict errors, and evaluate the impact of new information technologies and new procedures on flight crew performance. The human performance model component of these studies was developed using Air-MIDAS (Man Machine Integrated Design and Analysis System). The model was used to represent the flight deck crew responding to information systems and ATC.

The first study on surface operations represented flight crew responses to ground control commands for post landing roll out and taxi. The model used working memory limits, interference processes, and heuristics to successfully predict errors observed in the HITL simulation. The second round of studies concentrated on the use of SVS technologies to allow pilots to continue approach under visual minima. The perceptual model (visual sampling of information) was both statistically verified and validated against calibration data and HITL simulation data. With the validated model in place, we analyzed approach and go-around performance under standard and SVS technologies, and under conditions of approach and go-around decisions based on flight crew decisions and based on air traffic controller performance. The conclusions of this examination of SVS were that:

- SVS would not adversely affect the flight safety in approach, landing, and go-around phases regardless of the decision altitude and go-around triggers including the pilot-flying's intention at decision altitude and ATC's command, while it would allow approach and landing in conditions that would otherwise be unattainable;
- Small delays of action initiation in flight control were observed in the approach phase with SVS operations. This occurred because the chances of fixation on each display was decreased by adding the SVS to the conventional display configuration; and,
- No human performance degradation and no delay of task initiation were observed in the landing and go-around phases, although there were time shifts in the approach phase.

D-OMAR

S. Deutsch (Panelist) and R.W. Pew

The Distributed Operator Model Architecture (D-OMAR) provides an event-based simulator and a suite of representation language -- a frame language, a procedure language and a rule language -- that we have used to instantiate a cognitive architecture that is the basis for our aircrew and air traffic controller models. Using the procedural language, we have constructed a set of basic person

procedures -- perceptual, cognitive, and motor processes that are the building blocks for the expertise exhibited by aircrews and controllers.

As a general purpose simulator, D-OMAR has enabled us to also construct models for the essential elements of the commercial airspace: aircraft and their flight decks, ATC workplaces, airports with their runways, taxiways, and concourses, and the airways and navigation aides.

Our approach to examining aircrew error has been to build models that exhibit the robust behaviors of aircrews and then probe the models for the seams along which error can intrude (Deutsch & Pew, 2004). The infrequent errors in which aircrews mistakenly turned away for their designated concourse attracted our attention. Our analysis suggested that this was a point at which habit might intrude and lead to such an error. And indeed, one of the modeled competing sources for the action to take at an intersection was grounded in habit. Subsequent review of the human subject data validated the models prediction of the source of the error. In a similar manner, we modeled a case in which an aircrew, contrary to the ground controller's directive, turned toward their destination gate. To open the window for these errors, we constructed situations that prevented the first officer from prompting the captain on the correct turn to take at the intersection based on notes taken when the ground controller provided the taxi routing.

One aspect of our modeling of the use of the SVS, focused on our observation that it might be utilized as a second primary flight display (PFD) leading to an inefficient scan pattern. A review of the human subject data suggested that this might well be the case. To counter this potential problem, we designed a single attitude instrument combining PFD and SVS functionality. Model trials then predicted that the more efficient baseline scan pattern would be restored when using the combined PFD-SVS.

In each problem area addressed, the model and its underlying theory led to important new insights into the sources of aircrew behaviors.

IMPRINT / ACT-R

C. Lebiere (Panelist) and R. Archer

The approach that was used by our team to perform the approach and landing modeling task was an integration of the Improved Performance Research Integration Tool (IMPRINT) and the Adaptive Control of Thought – Rational (ACT-R) cognitive architecture. We found that a natural integration of the two tools had IMPRINT assuming the role of implementing the simulation and ACT-R assuming the role of the cognitive agents, i.e., the pilots. The IMPRINT model represents the state of the aircraft, its controls, and the environment. The ACT-R model represents the cognitive state of the pilots and their decision-making process. The two models communicate through a general, scalable, reusable interface called Link IMPRINT/ACT-R (LIA) that reduces the burden of integration from weeks or months of effort down to days. This approach of combining the strengths of task network and cognitive modeling for different portions of the same modeling scenario makes an important contribution to the capabilities of Human Performance Modeling: It allows complex simulations to be assembled in a modular fashion that makes explicit all communications requirements and enables high-fidelity cognitive modeling to be deployed where needed most in a tractable, affordable manner.

The cognitive models were built to emphasize the mechanisms and constraints of the cognitive architecture. Errors result from inherent limitations in the architecture's cognitive and perceptual abilities. Conversely, the model also exploits the architecture's powerful learning mechanisms to adapt to the introduction of new technology and other changes in its environment. We varied a number of parameters representing both variations in individual cognitive, perceptual and motor abilities as well as changes in the composition of the environment (e.g., by the addition of an SVS system) that affect the cognitive, perceptual and motor operations of the human operator (and cognitive model). This sensitivity analysis provides an indication of where the primary benefits of technological aids are likely to reside, as well as the most likely sources of error (Best, Lebiere, Schunk, Johnson, & Archer, 2004).

ADVANCES AND LESSONS LEARNED

K. Leiden (Panelist), D. C. Foyle and B. L. Hooey

Advances in Human Performance Modeling

Each of the HPMs discussed in this paper has extended their capabilities significantly to support the surface operations and SVS display modeling efforts. Considerable effort has been expended for the development of the external models to represent the aircraft flight dynamics, flight deck displays, and the communication link between external environment and HPM tool. (Some teams expended between 50-70% of their effort to represent these functions.) The ACT-R 5.0 and Air-MIDAS teams connected their models to higher-fidelity flight simulators and thus are poised to tackle future problems in which the closed-loop behavior between the pilot's action and the aircraft's response (or vice versa) is a key factor.

The usefulness of HPMs to the design and evaluation of new technology is determined to a significant extent by the core capabilities – visual attention allocation, workload, crew interactions, procedures, situation awareness, and error prediction. For example, the A-SA and ACT-R models focused specifically on what drives visual attention from a bottom-up (e.g., effort to move the eyes) as well as top-down perspective. Hence, if visual attention allocation needs to be understood for a particular technology, the ACT-R and A-SA modeling frameworks would more easily facilitate the analysis. In contrast, multiple operator models (e.g., pilot-flying and pilot-not-flying) are more easily accommodated by Air-MIDAS and D-OMAR. Thus, if flight crew or pilot/ATC interactions are expected to be significant drivers to a future modeling effort then the Air-MIDAS and D-OMAR frameworks would be more straightforward to apply. Of course, the respective HPMs and capabilities are dynamic. Each successive modeling effort in a complex environment such as aviation most likely adds to a framework's repertoire of capabilities.

Lessons Learned for the Aerospace Community

The modeling efforts revealed that HPMs, even those cognitive architectures that have traditionally been used in the context of psychological laboratory experiments, can indeed be useful tools for complex, context-dependent, domains such as aviation. Specifically, the tools can be used to address the design and evaluations of aviation displays, procedures, and operations.

Error Prediction and Mitigation. Across all of the modeling efforts, the tools were able to predict errors, or error vulnerabilities, that occurred because of situation awareness degradation, memory degradation and interference, airport layout, pilot expectation and habit, distraction, and workload. Further, it was shown that HPMs can be used to identify and evaluate the effectiveness of various technologies in the mitigation of such errors.

Display Design and Information Allocation. The models proved useful as tools to estimate the impact of new display technology on pilot scanning behavior. As such, HPMs can be used to inform display design and the allocation of information so as to optimize efficient scan patterns and increase the uptake of relevant information in a timely manner. As a salient example, the models

showed that when redundant information is overlaid on an SVS display (e.g., altitude, heading, speed on both traditional and SVS displays), pilots altered their baseline scan pattern to attend to the more easily acquired redundant information (based on saccade latency). Thus, the time spent attending to traditional displays under the SVS configuration was reduced significantly compared to the baseline configuration.

DISCUSSION

Under NASA's Human Performance Modeling Project, five models of human performance have been applied to a specific set of aviation problems. As a result, we are in a unique position to characterize some of the different ways in which specific models interact with the problem, and to note similarities and differences in how model representations affect the characterization of the modeling problem. The present panel discussion addresses this topic, and, in the near future, this will be documented as part of the HPM effort.

Some of the specific model characterizations include: 1) Model architecture and structures - To what extent do the specific architectures and structures in the various models impact: the user's choice of a modeling tool; the ability to describe/predict the data; and, the validation of results?; 2) Role of the external environment - How is the external environment captured in the model; how does the model interact with the external environment?; 3) Model predictive ability - To what extent do the specific modeling tools accurately model/predict behavior, produce emergent behavior, are predictive vs. simulation in nature, and, allow for the extrapolation to other non-tested display/procedural conditions?; and, 4) Usefulness/implications of the modeling results - What specific implications do the models make regarding procedures, communication, ATC - pilot interactions, intra-cockpit interactions, and, SVS display issues? Additionally, the future directions and challenges for human performance modeling in aviation are discussed and addressed.

ACKNOWLEDGMENTS

The work presented in this paper was supported by NASA's Aviation Safety and Security Program, Systemwide Accident Prevention, Human Performance Modeling element (21-078-20-10). For more information and for publication downloads see:

<http://human-factors.arc.nasa.gov/ih/hcsl/publications.html>

REFERENCES

- Best, B., Lebiere, C., Schunk, D., Johnson, I., and Archer, R. (2004). Validating a Cognitive Model of Approach Based on the ACT-R Architecture. (Technical Report). Micro Analysis and Design, Inc., Boulder, CO.
- Byrne, M. D., and Kirlik, A. (2005). Using computational cognitive modeling to diagnose possible sources of aviation error. *International Journal of Aviation Psychology*, vol. 15, pp. 135-155.
- Deutsch, S., and Pew, R. (2004). Examining new flight deck technology using human performance modeling. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA, pp. 108-112.
- Goodman, A., Hooey, B. L., Foyle, D. C., and Wilson, J. R. (2003). Characterizing Visual Performance During Approach and Landing With and Without a Synthetic Vision Display: A Part-Task Study. In D. C. Foyle, A. Goodman, and B. L. Hooey (Eds.), *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*, NASA Ames Research Center.
- Hooey, B. L., Foyle, D. C., and Andre, A. D. (2000). Integration of cockpit displays for surface operations: The final stage of a human-centered design approach. *SAE Transactions, Journal of Aerospace*, vol. 109, pp. 1053-1065.
- Leien, K., Keller, J. W., and French, J. W. (2001). Context of Human Error in Commercial Aviation. (Technical Report), Micro Analysis and Design, Inc., Boulder, CO.
- Leiden, K., Laughery, K. R., Keller, J. W., French, J. W., Warwick, W., and Wood, S. D. (2001). A Review of Human Performance Models for the Prediction of Human Error. (Technical Report), Micro Analysis and Design, Inc., Boulder, CO.
- Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., and Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors*, vol. 45, pp. 360-380.
- Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., and Zheng, S. (2005). Attention-Situation Awareness (A-SA) Model of Pilot Error. (Technical Report AHFD-04-15, NASA-04-5), University of Illinois at Urbana-Champaign, Institute of Aviation Human Factors Division, Savoy, Illinois.